METHOD AND SYSTEM FOR ANALYZING OPERATIONAL PARAMETER DATA FOR DIAGNOSTICS AND REPAIRS

[001] This application is continuing from U.S. Application Serial Number 09/688,105 filed October 13, 2000, which is a Continuation-In-Part of Application Serial Number 09/285,611 filed April 2, 1999. This application further claims the benefit of U.S. Provisional Application 60/162,045 filed October 28, 1999.

BACKGROUND OF THE INVENTION

[002] The present invention relates generally to machine diagnostics, and more specifically, to a system and method for processing historical repair data and operational parameter data for predicting one or more repairs from new operational parameter data from a malfunctioning machine.

[003] A machine such as locomotive includes elaborate controls and sensors that generate faults when anomalous operating conditions of the locomotive are encountered. Typically, a field engineer will look at a fault log and determine whether a repair is necessary.

[004] Approaches like neural networks, decision trees, etc., have been employed to learn over input data to provide prediction, classification, and function approximation capabilities in the context of diagnostics. Often, such approaches have required structured and relatively static and complete input data sets for learning, and have produced models that resist real-world interpretation.

[005] Another approach, Case Based Reasoning (CBR), is based on the observation that experiential knowledge (memory of past experiences - or cases) is applicable to problem solving as learning rules or behaviors. CBR relies on relatively little pre-processing of raw knowledge, focusing instead on indexing, retrieval, reuse, and archival of cases. In the diagnostic context, a case generally refers to a problem/solution description pair that represents a

diagnosis of a problem and an appropriate repair. More particularly, a case is a collection of fault log and corresponding operational and snapshot data patterns and other parameters and indicators associated with one specific repair event in the machine under consideration.

[006] CBR assumes cases described by a fixed, known number of descriptive attributes. Conventional CBR systems assume a corpus of fully valid or "gold standard" cases that new incoming cases can be matched against.

[007] U.S. Patent No. 5,463,768 discloses an approach which uses error log data and assumes predefined cases with each case associating an input error log to a verified, unique diagnosis of a problem. In particular, a plurality of historical error logs are grouped into case sets of common malfunctions. From the group of case sets, common patterns, i.e., consecutive rows or strings of data, are labeled as a block. Blocks are used to characterize fault contribution for new error logs that are received in a diagnostic unit.

[008] For a continuous fault code stream where any or all possible fault codes may occur from zero to any finite number of times and the fault codes may occur in any order, predefining the structure of a case is nearly impossible.

[009] U.S. Patent No. 6,343,236, assigned to the same assignee of the present invention, discloses a system and method for processing historical repair data and fault log data, which is not restricted to sequential occurrences of fault log entries and which provides weighted repair and distinct fault cluster combinations, to facilitate analysis of new fault log data from a malfunctioning machine. Further, U.S. Patent No. 6,415,395, assigned to the same assignee of the present invention, discloses a system and method for analyzing new fault log data from a malfunctioning machine in which the system and method are not restricted to sequential occurrences of fault log entries, and wherein the system and method predict one or more repair actions using

predetermined weighted repair and distinct fault cluster combinations. Additionally, U.S. Patent No. 6,336,065, assigned to the same assignee of the present invention, discloses a system and method that uses snapshot observations of operational parameters from the machine in combination with the fault log data in order to further enhance the predictive accuracy of the diagnostic algorithms used therein.

[010] It is believed that the inventions disclosed in the foregoing patent applications provide substantial advantages and advancements in the art of diagnostics. It would be desirable, however, to provide a system and method that uses anomaly definitions based on operational parameters to generate diagnostics and repair data. The anomaly definitions are different from faults in the sense that the information used can be taken in a relatively wide time window, whereas faults, or even fault data combined with snapshot data, are based on discrete behavior occurring at one instance in time. The anomaly definitions, however, may be advantageously analogized to virtual faults and thus such anomaly definitions can be learned using the same diagnostics algorithms that can be used for processing fault log data.

BRIEF DESCRIPTION OF THE INVENTION

[011] Generally, the present invention in one exemplary embodiment fulfills the foregoing needs by providing a method for analyzing operational parameter data from a malfunctioning locomotive or other large land-based, self-powered transport equipment. The method allows for receiving new operational parameter data comprising a plurality of anomaly definitions from the malfunctioning equipment. The method further allows for selecting a plurality of distinct anomaly definitions from the new operational parameter data. Respective generating steps allow for generating at least one distinct anomaly definition cluster from the plurality of distinct anomaly definitions and for generating a plurality of weighted repair and distinct anomaly definition cluster combinations. An identifying step allows for identifying at least one

repair for the at least one distinct anomaly definition cluster using the plurality of weighted repair and distinct anomaly definition cluster combinations.

[012] The present invention further fulfills the foregoing needs by providing in another aspect thereof a system for analyzing operational parameter data from a malfunctioning locomotive or other large land-based, self-powered transport equipment. The system includes a directed weight data storage unit adapted to store a plurality of weighted repair and distinct anomaly definition cluster combinations. A processor is adapted to receive new operational parameter data comprising a plurality of anomaly definitions from the malfunctioning equipment. Processor allows for selecting a plurality of distinct anomaly definitions from the new operational parameter data. Processor further allows for generating at least one distinct anomaly definition cluster from the selected plurality of distinct anomaly definitions and for generating a plurality of weighted repair and distinct anomaly definition cluster combinations. Processor 12 also allows for identifying at least one repair for the at least one distinct anomaly definition cluster using the plurality of predetermined weighted repair and distinct anomaly definition cluster combinations.

BRIEF DESCRIPTION OF THE DRAWINGS

[013] FIG. 1 is one embodiment of a block diagram of a system of the present invention for automatically processing repair data and operational parameter data from one or more machines and diagnosing a malfunctioning machine;

[014] FIG. 2 is an illustration of an exemplary data structure including data fields that may be used for specifying an anomaly definition and including exemplary new operational parameter data from a malfunctioning machine;

[015] FIG. 3 is a flowchart describing the steps for analyzing the new operational parameter data from a malfunctioning machine and predicting one or more possible repair actions;

[016] FIG. 4 is an illustration of distinct anomaly definitions identified in the new operational parameter data, such as may be represented in FIG. 2, and the number of occurrences thereof;

[017] FIGS. 5A-5D are illustrations of distinct fault anomaly definition clusters for the distinct faults identified in FIG. 4;

[018] FIG. 6 is a flowchart describing the steps for generating a plurality of predetermined cases, and predetermined repair and anomaly definition cluster combinations for each case;

[019] FIG. 7 is a flowchart describing the steps for determining predetermined weighted repair and anomaly definition cluster combinations;

[020] FIG. 8 is a printout of weighted repair and anomaly definition cluster combinations provided by the system shown in FIG. 1 for operational parameters that may be represented in FIG. 2, and a listing of recommended repairs;

[021] FIG. 9 is a flowchart further describing the step of predicting repairs from the weighted repair and anomaly definition cluster combinations shown in FIG. 8; and

[022] FIG. 10 is one embodiment of a flowchart describing the steps for automatically analyzing new operational parameter data from a malfunctioning machine and predicting one or more possible repair actions.

DETAILED DESCRIPTION OF THE INVENTION

[023] FIG. 1 diagrammatically illustrates one exemplary embodiment of a system 10 of the present invention. In one aspect, system 10 provides automated analysis of operational parameter data, from a malfunctioning machine such as a locomotive, and prediction of one or more possible repair actions.

[024] Although the present invention is described with reference to a locomotive, system 10 can be used in conjunction with any machine in which operation of the machine is monitored, such as a chemical, an electronic, a mechanical, a microprocessor machine and any other land-based, self-powered transport equipment.

[025] Exemplary system 10 includes a processor 12 such as a computer (e.g., UNIX workstation) having a hard drive, input devices such as a keyboard, a mouse, magnetic storage media (e.g., tape cartridges or disks), optical storage media (e.g., CD-ROMs), and output devices such as a display and a printer. Processor 12 is operably connected to a repair data storage unit 20, an operational parameter data storage unit 22, a case data storage unit 24, and a directed weight data storage unit 26.

[026] FIG. 2 shows an exemplary data structure 50 comprising a plurality of data fields, generally associated with anomaly definitions based on operational parameter data. As shown in FIG. 2, a set of data fields 52 may include general information regarding each anomaly definition, such as anomaly definition identifier, objective, explanatory remarks, message to be automatically generated upon detection of a respective anomaly definition, personnel responsible for handling a respective anomaly definition, locomotive model and configuration, etc. As further shown in FIG. 2, a set of data fields 54 may include observations indicative of locomotive operating conditions that may be associated with an anomaly definition, including statistics data and trend data that may be extracted from such observations. FIG. 2 further shows a set of data fields 56 that may include operational operational parameter data that may be associated with a given anomaly definition. For example, if parameter 1 is outside a predefined range, and the standard deviation of parameter 2 is beyond a predefined level, and parameter 3 exhibits a trend that exceeds a predefined rate of change, and parameter 4 is outside another predefined range under a given set of locomotive operating condition, then, assuming each of the above conditions is met, and further assuming that there is an anomaly definition specifying

each of such conditions, that would constitute detection of such anomaly definition, that is, the occurrence of each of such events would trigger that anomaly definition. It will be appreciated that the level of information that can be obtained from anomaly definitions based on operational parameter data comprising a selectable time window is more statistically robust compared to fault log data that are based on the occurrence of single instance events. The inventors of the present invention have advantageously recognized that diagnostics algorithm techniques typically associated with the processing of fault log data may now be extended to processing anomaly definitions based on continuous operational parameter data, as opposed to singular time events. As used herein operational parameter data refers to continuous or non-discrete data. That is, data that may be expressed in numerical ranges such as engine speed, voltages, etc., or data that may be monitored over a desired time window for trends, shifts, changes, etc., as opposed to data indicative of discrete states. Of course, the term continuous data does not exclude digitally sampled data since such data may be observed over a desired time window provided the sampling rate is sufficiently fast relative to the time window to detect trends, shifts, changes, etc.

[027] FIG. 3 is a flowchart which generally describes the steps for analyzing new operational parameter data 200 (FIG. 1). As shown in FIG. 3 at 232, the new operational parameter data comprising a plurality of anomaly definitions from a malfunctioning machine is received. At 233, a plurality of distinct anomaly definitions from the new operational parameter data is identified, and at 234, the number of times each distinct anomaly definition occurred in the new operational parameter data is determined. As used herein, the term "distinct anomaly definition" is an anomaly definition or anomaly code which differs from other anomaly definitions or anomaly codes so that, as described in greater detail below, if the operational parameter data includes more than one occurrence of the same anomaly definition or anomaly code, then similar anomaly definitions or anomaly codes are identified only once. As will become apparent from the discussion below, in one exemplary embodiment, it

is the selection or triggering of distinct anomaly definitions which is important and not the order or sequence of their arrangement.

[028] FIG. 4 shows an exemplary plurality of distinct anomaly definitions and the number of times in which each distinct anomaly definition occurred for operational parameter 220 (FIG. 2). In this example, anomaly definition code 7311 represents a phase module malfunction which occurred 24 times, anomaly definition code 728F indicates an inverter propulsion malfunction which occurred twice, anomaly definition code 76D5 indicates an anomaly definition which occurred once, and anomaly definition code 720F indicates an inverter propulsion malfunction which occurred once.

[029] With reference again to FIG. 3, a plurality of anomaly definition clusters is generated for the distinct anomaly definitions at 236. FIGS. 5A-5D illustrate the distinct anomaly definition clusters generated from the distinct anomaly definitions extracted from operational parameter data 200. Four single anomaly definition clusters (e.g., anomaly definition code 7311, anomaly definition code 728F, anomaly definition code 76D5, and anomaly definition code 720F) are illustrated in FIG. 5A. Six double anomaly definition clusters (e.g., anomaly definition codes 76D5 and 7311, anomaly definition codes 76D5 and 728F, anomaly definition codes 76D5 and 720F, anomaly definition codes 7311 and 728F, anomaly definition codes 7311 and 720F, and anomaly definition codes 728F and 720F) are illustrated in FIG. 5B. Four triple anomaly definition clusters (e.g., anomaly definition codes 76D5, 7311, and 728F), anomaly definition codes 76D5, 7311, and 720F, anomaly definition codes 76D5, 728F, and 720F, and anomaly definition codes 7311, 728F, and 720F) are illustrated in FIG. 5C, and one quadruple anomaly definition cluster (e.g., 76D5, 7311, 728F, and 720F) is illustrated in FIG. 5D.

[030] From the present description, it will be appreciated by those skilled in the art that an anomaly definition log having a greater number of distinct anomaly definitions would result in a greater number of distinct anomaly definition clusters (e.g., ones, twos, threes, fours, fives, etc.).

[031] At 238, at least one repair is predicted for the plurality of anomaly definition clusters using a plurality of predetermined weighted repair and anomaly definition cluster combinations. The plurality of predetermined weighted repair and anomaly definition cluster combinations may be generated as follows.

[032] With reference again to FIG. 1, processor 12 is desirably operable to process historical repair data contained in a repair data storage unit 20 and historical operational parameter data contained in an operational parameter data storage unit 22 regarding one or more locomotives.

[033] For example, repair data storage unit 20 includes repair data or records regarding a plurality of related and unrelated repairs for one or more locomotives. Operational parameter data storage unit 22 includes operational parameter data or records regarding a plurality of anomaly definitions occurring for one or more locomotives.

[034] FIG. 6 is a flowchart of an exemplary process 50 of the present invention for selecting or extracting repair data from repair data storage unit 20 and operational parameter data from the operational parameter data storage unit 22 and generating a plurality of cases, and repair and anomaly definition cluster combinations.

[035] Exemplary process 50 comprises, at 52, selecting or extracting a repair from repair data storage unit 20 (FIG. 1). Given the identification of a repair, the present invention searches operational parameter data storage unit 22 (FIG. 1) to select or extract anomaly definitions occurring over a predetermined period of time prior to the repair, at 54. At 56, the number of times each distinct anomaly definition occurred during the period of time is determined.

[036] A repair and corresponding distinct anomaly definitions are summarized and stored as a case, at 60. For each case, a plurality of repair

and anomaly definition cluster combinations are generated at 62 (in a similar manner as described for the new operational parameter data).

[037] Process 50 is repeated by selecting another repair entry from repair data to generate another case, and to generate a plurality of repair and anomaly definition cluster combinations. Case data storage unit 24 desirably comprises a plurality of cases comprising related and unrelated repairs.

[038] FIG. 7 is a flowchart of an exemplary process 100 of the present invention for generating weighted repair and anomaly definition cluster combinations based on the plurality of cases generated in process 50. Process 100 comprises, at 101, selecting a repair and anomaly definition cluster combination, and determining, at 102, the number of times the combination occurs for related repairs. The number of times the combination occurs in the plurality of cases of related and unrelated repairs, e.g., all repairs for similar locomotives, is determined at 104. A weight is determined at 108 for the repair and distinct anomaly definition cluster combination by dividing the number of times the distinct anomaly definition cluster occurs in related cases by the number of times the distinct anomaly definition cluster occurs in all, e.g., related and unrelated cases, and the weighted repair and distinct anomaly definition cluster combination is desirably stored in a directed weight data storage unit 26.

[039] FIG. 8 illustrates an exemplary printout 250 of the results generated by system 10 (FIG. 1) based on operational parameter data 200 (FIG. 1), in which in a first portion 252, a plurality of corresponding repairs 253, assigned weights 254, and anomaly definition clusters 255 are presented. As shown in a second portion 260 of printout 250, five recommendations for likely repairs actions are presented for review by a field engineer.

[040] FIG. 9 is a flowchart of an exemplary process 300 for determining and presenting the top most likely repair candidates which may include repairs derived from predetermined weighted repair and distinct anomaly definition

cluster combinations having the greatest assigned weighted values or repairs which are determined by adding together the assigned weighted values for anomaly definition clusters for related repairs.

[041] As shown in FIG. 9, initially, a distinct anomaly definition cluster generated from the new operational parameter data is selected at 302. At 304, predetermined repair(s) and assigned weight(s) corresponding to the distinct anomaly definition cluster are selected from directed weight storage unit 26 (FIG. 1).

[042] At 306, if the assigned weight for the predetermined weighted repair and anomaly definition cluster combination is determined by a plurality of cases for related and unrelated repairs which number is less than a predetermined number, e.g., 5, the cluster is excluded and the next distinct anomaly definition cluster is selected at 302. This prevents weighted repair and anomaly definition cluster combinations which are determined from only a few cases from having the same effect in the prediction of repairs as weighted repair and anomaly definition cluster combinations determined from many cases.

[043] If the number of cases is greater than the predetermined minimum number of cases, at 308, a determination is made as to whether the assigned value is greater than a threshold value, e.g., 0.70 or 70%. If so, the repair is displayed at 310. If the anomaly definition cluster is not the last anomaly definition cluster to be analyzed at 322, the next distinct anomaly definition cluster is selected at 302 and the process is repeated.

[044] If the assigned weight for the predetermined weighted repair and anomaly definition cluster combination is less than the predetermined threshold value, the assigned weights for related repairs are added together at 320. Desirably, up to a maximum number of assigned weights, e.g., 5, are used and added together. After selecting and analyzing the distinct anomaly definition clusters generated from the new operational parameter data, the

repairs having the highest added assigned weights for anomaly definition clusters for related repairs are displayed at 324.

[045] With reference again to FIG. 8, repairs corresponding to the weighted repair and anomaly definition cluster combinations in which the assigned weights are greater than the threshold value are presented first. As shown in FIG. 8, repair codes 1766 and 1777 and distinct anomaly definition cluster combinations 7311, 728F, and 720F, have an assigned weight of 85% and indicate a recommended replacement of the EFI.

[046] As also shown in FIG. 8, repairs for various anomaly definition clusters having the highest added or total weight are presented next. For example, repair code 1677 which corresponds to a traction problem has a totaled assigned weight of 1.031, repair code 1745 which corresponds to a locomotive software problem has a totaled assigned weight of 0.943, and repair code 2323 which corresponds to an overheated engine has a totaled assigned weight of 0.591.

[047] Advantageously, the top five most likely repair actions are determined and presented for review by a field engineer. For example, up to five repairs having the greatest assigned weights over the threshold value are presented. When there is less than five repairs which satisfy the threshold, the remainder of recommended repairs are presented based on a total assigned weight.

[048] Desirably the new operational parameter data is initially compared to a prior operational parameter data from the malfunctioning locomotive. This allows determination whether there is a change in the operational parameter data over time. For example, if there is no change, e.g., no new anomaly definitions, then it may not be necessary to process the new operational parameter data further.

[049] FIG. 10 illustrates a flowchart of an exemplary automated process 500 for analyzing operational parameter data from a locomotive, e.g., new operational parameter data which is generated every day, using system 10.

In particular, process 500 accommodates the situation where a prior repair is undertaken or a prior repair is recommended within the predetermined period of time over which the operational parameter data is analyzed. This avoids recommending the same repair which has been previously recommended and/or repaired.

[050] At 502, new operational parameter data is received which includes anomaly definitions occurring over a predetermined period of time, e.g., 14 days. The operational parameter data is analyzed, for example as described above, generating distinct anomaly definition clusters and comparing the generated anomaly definition clusters to predetermined weighted repair and anomaly definition cluster combinations.

[051] At 504, the analysis process may use a thresholding process described above to determine whether any repairs are recommended (e.g., having a weighted value over 70%). If no repairs are recommended, the process is ended at 506. The process is desirably repeated again with a download of new operational parameter data the next day.

[052] If a repair recommendation is made, existing closed (e.g., performed or completed repairs) or prior recommended repairs which have occurred within the predetermined period of time are determined at 508. For example, existing closed or prior recommended repairs may be stored and retrieved from repair data storage unit 20. If there are no existing or recommended repairs than all the recommended repairs at 504 are listed in a repair list at 700.

[053] If there are existing closed or prior recommended repairs, then at 600, any repairs not in the existing closed or prior recommended repairs are listed in the repair list at 700.

[054] For repairs which are in the existing closed or prior recommended repairs, at 602, the look-back period (e.g., the number of days over which the anomaly definitions are chosen) is revised. Using the modified look-back or

shortened period of time, the modified operational parameter data is analyzed at 604, as described above, using distinct anomaly definition clusters, and comparing the generated anomaly definition clusters to predetermined weighted repair and anomaly definition cluster combinations.

[055] At 606, the analysis process may use the thresholding process described above to determine whether any repairs are recommended (e.g., having a weighted value over 70%). If no repairs are recommended, the process is ended at 608 until the process is stated again with a new operational parameter data from the next day, or if a repair is recommended it is added to the repair list at 700.

[056] From the present description, it will be appreciated by those skilled in the art that other processes and methods, e.g., different thresholding values or operational parameter data analysis which does not use distinct anomaly definition clusters, may be employed in predicting repairs from the new operational parameter data according to process 500 which takes into account prior performed repairs or prior recommended repairs.

[057] Thus, the present invention provides in one aspect a method and system for processing a new operational parameter which is not restricted to sequential occurrences of anomaly definitions or error log entries. In another aspect, the calibration of the diagnostic significance of anomaly definition clusters is based upon cases of related repairs and cases for all the repairs.

[058] While the invention has been described with reference to preferred embodiments, it will be understood by those skilled in the art that various changes may be made and equivalents may be substituted for elements thereof without departing from the scope of the invention. In addition, many modifications may be made to adapt a particular situation or material to the teachings of the invention without departing from the essential scope thereof. Therefore, it is intended that the invention not be limited to the particular

embodiments disclosed herein, but that the invention will include all embodiments falling within the scope of the appended claims.